n a c l o

The Ninth Annual

North American Computational Linguistics Olympiad

2015

www.naclo.cs.cmu.edu

SOLUTIONS
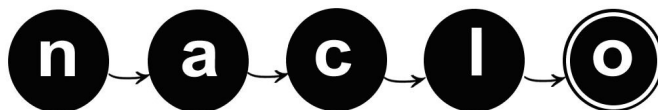## Invitational Round
March 12, 2015

(I) Old English (1/1)

I1.
    a.   The king loved you all
    b.   You all loved that girl
    c.   We two loved you two

I2.
    a.   se æþeling lufode þæt cild
    b.   þæt cild lufode þone æþeling
    c.   we lufodon þæt cild
    d.   þæt cild inc lufode

(J) Georgian Transitive Verbs (1/1)

Simplified conjugation table (for the purposes of this problem)

	Present tense	Aorist tense
1ˢᵗ sing.	ვ- + verb root [VR]	Preverb vowel(PV)+ -ვ- + VR
2ⁿᵈ sing.	VR	PV+VR
3ʳᵈ sing.	VR + -ს	PV+ VR + -a
1ˢᵗ pl.	ვ- + VR + -თ	PV + -ვ- +VR + -თ
2ⁿᵈ pl.	VR + -თ	PV + VR + -თ
3ʳᵈ pl.	VR + -ენ	PV+ VR + -ს

Present Tense

	To open[1]	To build	To receive
1ˢᵗ sing.	ვაღებ "vagheb"	ვაშენებ "vasheneb"	ვიღებ vigheb
2ⁿᵈ sing.	აღებ "agheb"	აშენებ "asheneb"	იღებ igheb
3ʳᵈ sing.	აღებს "aghebs"	აშენებს "ashenebs"	იღებს ighebs
1ˢᵗ pl.	ვაღებთ "vaghebt"	ვაშენებთ "vashenebt"	ვიღებთ vighebt
2ⁿᵈ pl.	აღებთ "aghebt"	აშენებთ "ashenebt"	იღებთ ighebt
3ʳᵈ pl.	აღებენ "agheben"	აშენებენ "asheneben"	იღებემ igheben

Aorist Tense

|	To make	To repair	To write
1ˢᵗ sing.	გავაკეთე "Gavakete"	შევაკეთე "sevakete"	დავწერე "davtsere"
2ⁿᵈ sing.	გააკეთე "gaakete"	შეაკეთე "sheakete"	დაწერე "datsere"
3ʳᵈ sing.	გააკეთა "gaaketa"	შეაკეთა "sheaketa"	დაწერა "datsera"
1ˢᵗ pl.	გააკეთეთ "gavaketet"	შევაკეთეთ "shevaketet"	დავწერეთ "davtseret"
2ⁿᵈ pl.	გააკეთეთ "gaaketet"	შეაკეთეთ "sheaketet"	დაწერეთ "datseret"
3ʳᵈ pl.	გააკეთეს "gaaketes"	შეაკეთეს "sheaketes"	დაწერეს "datseres"

J1.
   a. They are opening the letter
   b. You (pl.) have repaired the computer.
   c. Adam is building a museum.
   d. We have written a story/history.

J2.

   a. ფარდულს აღებ.
   b. აკტომობილი შევაკეთე.
   c. სადიღს ვიღებთ.
   d. საჩუქარი გააკეთეს.

[1]Georgian verbs do not have an infinitive.

n → a → c → l → o
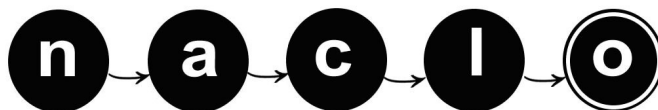
(K) The Dualization Game (1/2)

K1. Any three nouns that end in x, s, or z and take an -es plural ending – ideally one each for each of those three last letters. For example, buzz ⇒ buzzes, class ⇒ classes, tax ⇒ taxes.

K2. * quiz ⇒ quizes, * child ⇒ childs

K3. * nidaahné ⇒ niahné, * nidaniiché ⇒ niniiché

K4. Complete Dualizer Machine 2.0 (cells not given in question are bolded):

Entry state	Read	Write	Move	Exit State
S0	d	Ø	R	S1
S0	[otherwise]	[N/A]	R	S0
S1	[otherwise]	**Ø**	**L**	S2
S2	**[otherwise] or Ø**	[N/A]	**L**	S3
S3	i	[N/A]	**L**	S4
S3	**[otherwise]**	**[N/A]**	[N/A]	HALT
S4	**n**	[N/A]	**R**	S5
S4	[otherwise]	[N/A]	[N/A]	HALT
S5	[otherwise]	**a**	[N/A]	HALT

K5. Complete Second-Personizer Machine (cells not given are bolded):

Entry state	Read	Write	Move	Exit State
S0	**s**	**ł or others[1]**	**R**	S1
S0	[otherwise]	[N/A]	R	S0
S1	**h**	**Ø**	**L**	S2
S1	[otherwise]	**[N/A]**	L	S2
S2	[otherwise]	**[N/A] or ł**	**L**	S3
S3	**í**	[N/A]	[N/A]	HALT
S3	[otherwise]	**[N/A] or i**	L	S4
S4	**i, í**	[N/A]	[N/A]	HALT
S4	**y**	**n**	[N/A]	HALT
S4	[otherwise]	[N/A]	**R**	S5
S5	[otherwise]	**í**	[N/A]	HALT

[1]Any other letter, [N/A], and Ø are all acceptable answers here – however, if an answer other than ł is given, the Write column of S2 must be ł. If ł is given in S0 Write, either answer is acceptable for S2 Write.

n → a → c → l → o
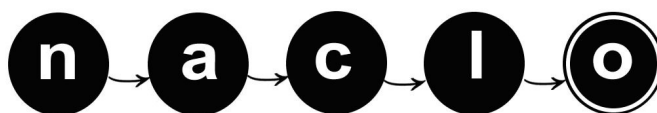
K6. íísínísts'ą́ą́' will be transformed unsuccessfully. Both verbs have predictable second-person forms (íísíníłts'ą́ą́' and bigháníłdééh respectively), but the machine will attempt to apply its transformations on and around the first s in íísínísts'ą́ą́', rather than the second s. The unsuccessful output for this verb will be ííłínísts'ą́ą́'.

L1.
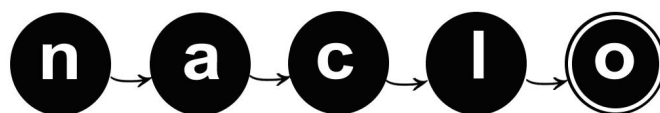
a. fito
b. valo amby enimpolo sy sivinjato sy dimy arivo sy alina
c. telo amby fitopolo sy dimanjato sy sivy arivo sy sivy alina
d. valo amby telopolo sy eninjato sy valo alina
e. fito ambin'ny folo sy valonjato

The solved crossnumber puzzle:

¹7	²7	■	³1	⁴7
⁵9	0	⁶3	6	4
⁷7	1	2	1	5
⁸1	5	9	8	6
⁹1	0	■	¹⁰7	1

Further Explanation
This chart shows how to say all pertinent place values in Malagasy:

Digit	x 1	x 10	x 100	x 1,000	x 10,000
1	iray/iraika	folo	zato	arivo	alina
2	roa	roapolo	roanjato	roa arivo	roa alina
3	telo	telopolo	telonjato	telo arivo	telo alina
4	efatra	efapolo	efajato	efatra arivo	afatra alina
5	dimy	dimampolo	dimanjato	dimy arivo	dimy alina
6	enina	enimpolo	eninjato	enina arivo	enina alina
7	fito	fitopolo	fitonjato	fito arivo	fito alina
8	valo	valopolo	valonjato	valo arivo	valo alina
9	sivy	sivifolo	sivinjato	sivy arivo	sivy arivo

Other notes

- Digits are written from left to right with the digit of lowest magnitude coming first.
- Between each digit is a connecting word, chosen as follows:
  - If the word is connecting the ones place and the tens place, and if the tens place is 1, then the word used is *ambin'ny*.
  - If the word is connecting the ones place and the tens place, and if the tens place is not 1, then the word used is *amby*.
  - In all other cases, *sy* is used.
- 1 is *iray* on its own but *iraika* when it is the ones digit of a larger number.
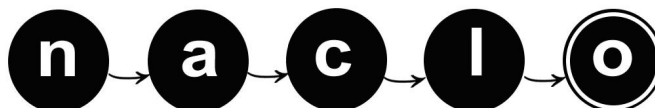
Simple Path to the Solution

(1) Note that the answer to Sivy-Across is *folo*, which is also the number for a different across answer. Because all answers are at least 2 digits long, and because 10 is the only answer number that is not a single digit, *folo* must mean 10. Since this is 2 digits long, Sivy-Across has to be 1-Across, 3-Across, 9-Across, or 10-Across. It cannot be 1- or 3- Across because that would mean that 2- or 4-Down would start with a zero (which is not allowed, as stated in the directions), and it cannot be 10-Across because we know that *folo* (not *sivy*) means 10. Therefore, *sivy* must mean 9, and we can fill in that answer in the grid:

1	2	■	3	4
5		6		
7				
8				
9 1	0	■	10	

(2) Notice that two numbers, 1 and 3, appear as the numbers for both an Across and a Down answer. *Iray* and *telo* are the only Malagasy answer numbers that appear in both Across and Down, so they must correspond to 1 and 3 (though we do not know yet which is which). The remaining three Across numbers (5, 7, and 8) must correspond to *dimy*, *fito*, and *valo*, in some unknown order.

(3) Notice that every answer follows the form [Word that is also an answer number] [*amby* or *ambin'ny*] [*folo* or word ending in -*polo*] [*sy*] [*zato* or word ending in -*jato*] [*sy*] [*arivo* or two words, the second of which is *arivo*] [*sy*] [*alina* or two words, the second of which is *alina*]. Note that not all answers include all slots, but when present the slots are always in this order. From this pattern, it is likely that *amby/ambin'ny* and *sy* act as connectors, probably meaning "and," and that the other five slots correspond to the (at most) five digits present in the answers.

(4) The next big discovery that the solver must make is that the digits are listed in reverse order--i.e., 1,234 is spelled out as "four and thirty and two hundred and one thousand." There are many ways to figure this out, but here is one example of how to do so: Notice that, from the number template in (3), the last few slots are always null in any two-digit numbers. That is, no two-digit answers have anything in the [*zato* or word ending in -*jato*] slot or the [*alina* or two words, the second of which is *alina*] slot or the [*arivo* or two words, the second of which is *arivo*] slot. This means that the later slots probably stand for the higher digits because those are the digits that are not present in 2-digit numbers. It also makes sense that the first slot [Word that is also an answer number] is the ones digit, since the answer numbers (except for 10) are all the single digits; in addition, we know that *folo* means ten, so it makes sense that the second slot [*folo* or word ending in -*polo*] is the tens place. Thus, we can conclude that the digits are listed from lowest power of ten to highest. This gives us the following correspondences between slots and powers of 10:
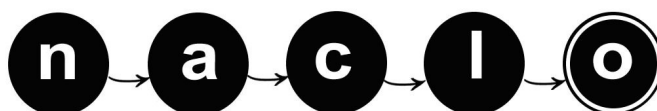
1: [Word that is also an answer number]
10: [*folo* or word ending in *-polo*]
100: [*zato* or word ending in *-jato*]
1,000: [*arivo* or two words, the second of which is *arivo*]
10,000: [*alina* or two words, the second of which is *alina*]

(5) Remember from (2) that *iray* and *telo* stand for 1 and 3, not necessarily in that order. We already know that the last digit of 1-Down is 1 (because we filled in 9-Across as 10). Since 1 is the last digit in the numeral, we know it will be the first digit listed in the spelled-out number. The first digits of Iray-Down and Telo-Down are *iraika* and *fito*, respectively, so one of those must mean 1. Since we know that either *telo* or *iray* means 1 and that either *iraika* or *fito* means 1, it makes the most sense to say that the similarly spelled *iray* and *iraika* are two forms of 1. This also means that *telo* is 3.

(6) 10-Across is *iraika amby fitopolo* which means that its ones digit is 1. 4-Down has to have the same ones digit as 10-Across, and since Efatra-Down is the only down answer starting with *iraika* (other than Iray-Down, which we already know is 1-Down), *efatra* must mean 4. We can now also fill in that 1 in the grid:

1	2		3	4
5		6		
7				
8				
9 1	0		10	1

(7) The two Down numbers left to identify are 6-Down (which is 3 digits long) and 2-Down (which is 5 digits long), which must match up to Enina-Down and Roa-Down. Both Enina- and Roa-Down only seem to list 3 digits. However, Roa (unlike Enina) includes an *alina* digit, which from the template we know to be the ten thousands digit. Therefore, Roa must be the 5-Digit one. This means *roa* is 2 and *enina* is 6. We can now fill in the ones digit of Enina-Down, since we know *sivy* means 9.

1	2		3	4
5		6		
7				
8		9		
9 1	0		10	1

(8) Valo-Across has a hundreds digit of *sivinjato*, and no other Across answers have this same hundreds digit. Since *sivy* means 9, *sivinjato* is probably 900. With the 9 we just added in (7), this tells us that *valo* means 8.

n → a → c → l → o

The hundreds digit of Dimy-Across (*telonjato*) matches the hundreds digit of 6-Down, and 5-Across has the same hundreds digit as 6-Down, so *dimy* must mean 5. This leaves *fito* to mean 7. We now know all the digits from 1 to 10:

- 1 = iray/iraika
- 2 = roa
- 3 = telo
- 4 = efatra
- 5 = dimy
- 6 = enina
- 7 = fito
- 8 = valo
- 9 = sivy
- 10 = folo

(9) We can now fill in all the ones digits of all the answers:

1	2 7	■	3	4 7
5		6		4
7				5
8		9		8
9 1	0	■	10 7	1

(10) Now 4-Down is completely filled in and can be used to understand the higher-valued digits (if the solver didn't figure this out already). The tens digit is *folo* (for ten) or, for a number (k * 10), it is the word corresponding to the digit k followed by the suffix *-polo* (possibly with some phonological changes at morpheme boundaries). The hundreds digit is similar, except with *zato* as the base word (meaning 100) which becomes the suffix *-jato*. Lastly, the thousands and ten thousands place are denoted by [number] *arivo* or [number] *alina*, where [number] is null if the digit is one but is the name of that single digit otherwise. At this point, it is also worth noticing that the connector *ambin'ny* is used between the ones and tens places if the tens place is 1; the connector *amby* is used between the ones and tens places otherwise; and the connector *sy* is used between all other decimal places. Using this, we can fill in the rest of the grid:

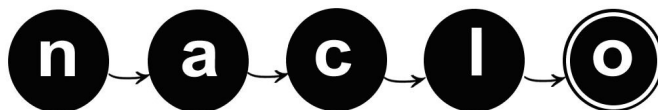| 1 7	2 7	■	3 1	4 7
5 9	0	6 3	6	4
7 7	1	2	1	5
8 1	5	9	6	8
9 1	0	■	10 7	1

(11) Answering the questions:
- *a* and *b* are both numbers taken from the grid; thus, we can already answer those immediately.
- *c* is a new number, but each separate digit has been seen before: 90,000 is *sivy alina*, 9,000 is *sivy arivo*, 500 is *dimanjato*, 70 is *fitopolo*, and 3 is *telo*. Thus, this overall number is *telo amby fitopolo sy dimanjato sy sivy arivo sy sivy alina*
- For *d* and *e*, there are values we have not seen yet. It is helpful to refer to this table of the values we have seen:

Digit	× 1	× 10	× 100	× 1,000	× 10,000
1	iray/iraika	folo	zato	arivo	alina
2	roa	roapolo	roanjato		
3	telo		telonjato		
4	efatra			efatra arivo	
5	dimy	dimampolo	dimanjato	dimy arivo	
6	enina	enimpolo		enina arivo	
7	fito	fitopolo	fitonjato		fito alina
8	valo	valopolo			
9	sivy		sivinjato	sivy arivo	sivy arivo

- In *d*, we have 80,000; 600; and 30 as unfamiliar digits. The ten thousands place is formed very regularly as [digit] *alina*, so it is easy to see that 80,000 is *valo alina*. 600 is trickier: we have to notice that 50 and 500 are *dimampolo* and *dimanjato*, respectively, and therefore the fact that 60 is *enimpolo* means that 600 is likely *eninjato* (rather than, say, *eninanjato*). Lastly, since digits written as 3 seem to follow the same pattern as digits written as 7 or 8, it is most likely that 30 is *telopolo*. This gives a final answer of *valo amby telopolo sy eninjato sy valo alina*.
- In *e*, 800 is new, but comparison with 700 and 300 give us *valonjato* as a likely representation of it. Remembering from (10) that *ambin'ny* is used as the connector when forming a number in the teens, we get *fito ambin'ny folo sy valonjato*.

M1. ("" (+H +a +u +s ->
"Haus" (+e +s ->
"Hauses"(-a +ä -s +r ->
"Häuser"(+n ->
"Häusern")))))   Cost: 11
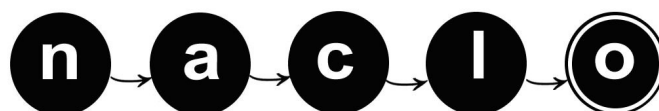
M2. ("" (+H +a +u +s ->
"Haus" (+e ->
"Hause" (+s ->
"Hauses")
(-a +ä +r ->
"Häuser"(+n ->
"Häusern")))))   Cost: 10

M3. 7

("" (+H +a +u +s +e ->
"Hause" (-e ->
"Haus")
(+s ->
"Hauses")
(-a +ä +r ->
"Häuser"(+n ->
"Häusern"))))   Cost: 11

("" (+H +a +u +s ->
"Haus" (+e ->
"Hause" (+s ->
"Hauses")
(-a +ä +r +n ->
"Häusern" (-n ->
"Häuser")))))   Cost: 11

("" (+H +a +u +s ->
"Haus" (+e ->
"Hause" (+s ->
"Hauses"(-a +ä -s +r ->
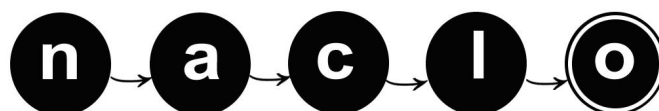"Häuser"(+n ->
"Häusern")))))   Cost: 11

n a c l o

```
("" (+H +a  +u  +s   ->
   "Haus"   (+e +s   ->
       "Hauses")
            (+e ->
       "Hause ")
            (-a  +ä   +r   ->
       "Häuser "(+n ->
            "Häusern"))))          Cost: 11

("" (+H +a  +u  +s   ->
   "Haus"   (+e +s   ->
       "Hauses"(-s   ->
            "Hause ")
                (-a  +ä   +r   ->
            "Häuser "(+n  ->
                "Häusern")))))     Cost: 11

("" (+H +a  +u  +s   ->
   "Haus"   (-a  +ä   +e   +r   ->
       "Häuser "(+n  ->
            "Häusern"))
         (+e   ->
            "Hause " (+s  ->
                "Hauses"))))       Cost: 11
```

The seventh MST is the cost 10 tree from **M2**.

M4. Haus is encoded from Häusern and all other words are encoded from "", for a cost of 5 + 5 + 6 + 6 + 7 = 29

(N) Maxakalí (1/1)

1. ka'õgãhã
2. ka'ok
3. kuxa
4. kuxa ka'ok
5. mĩkax
6. mĩkaxxax
7. mĩptut
8. mĩptut mõg
9. mĩptut mõg kuxa
10. mĩptut mõg pata
11. mõg
12. mõgãhã
13. pa
14. pa ka'ok
15. pata
16. pataxax
17. paxax
18. xax

G. harden
F. hard
H. heart
P. stubborn
J. knife
K. knife sheath
I. house
N. motor vehicle
M. motor
Q. to tire
E. go
L. lead
B. eye
R. wide awake
D. foot
O. shoe
C. eyelid
A. cover

O1.

|			
1.	Q	17.	J
2.	O	18.	CC
3.	P	19.	S
4.	T	20.	AA
5.	B	21.	M
6.	EE	22.	X
7.	H	23.	Y
8.	W	24.	DD
9.	D	25.	I
10.	N	26.	E
11.	K	27.	G
12.	R	28.	C
13.	F	29.	V
14.	Z	30.	L
15.	BB	31.	A
16.	U		

O2.
   a. neeg Hmoob neeg Nplog
   b. ua liab ua cuam
   c. muaj nyiaj muaj kub

Explanation

Solving this problem involves discovering a number of generalizations about both the meaning and the formal structure of the expressions that are given in both Hmong and English. These generalizations are sufficient to allow the solver to translate the English expressions under translation. However, as will be seen, one of these generalizations is very subtle and only the most careful linguists are likely to notice it.

One source of evidence that is likely to help investigators in the initial stages is the use of capitalization with proper nouns even in non-initial positions. One of these words, *Hmoob*, is already known to translate as 'Hmong' based on the introductory text. The remaining word, *Nplog*, has to be 'Lao'. This means that *hais lus Hmong* must translate as 'speak Hmong language', and *Neeg Nplog pe mlom* must translate as 'Lao people reverse images'. These matches with provide crucial information regarding the meaning of Hmong words. Another, related, source of information has to do with capitalization and punctuation: full sentences are
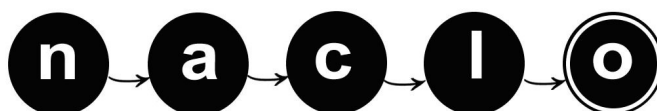
capitalized and terminated with periods while phrases that do not form a complete sentence are not. This allows the investigator to divide both the Hmong and English sentences into these two categories and thus makes it easier to find matches.

These constraints can help the analyst find the setting of two parameters in Hmong: the order to head-modifier constructions (like 'Lao people') and the word order of whole clauses (subject-verb-object, subject-object-verb, etc.). These questions can be resolved, in part, by examining the distribution of words in items that are already known. For example, *neeg* occurs in 5 items and 'people' also occurs in five items. This is consistent with the hypothesis that the *neeg* in *neeg nplog* means 'people' and that modifiers follow the nouns that they modify in Hmong. If this is the case, it appears that the subject occurs at the beginning of the sentence in Hmong. It remains to be determined whether the direct object occurs before or after the verb. There are two words left unaccounted for in the expression *Neeg Nplog þe mlom*, namely *þe* and *mlom*. Since these translate the sentence 'Lao people reverence images,' on of these words is likely to mean 'reverence' (the verb) and the other is likely to mean 'images' (the object). It is straightforward to note that the expression *þe mlom* also occurs in the phrase *þe dab þe mlom*.

There is only one other item that includes the sense 'reverence images' and this is 'reverence images and spirits'. This must match *þe dab þe mlom*. This item has a curious structure: it has an ABAC pattern. What could the repeated word mean? There are multiple ways of reaching a conclusion on this subject. One helpful approach involves the semantics of 'reverence images and spirits'. In this phrase, 'reverence' applies to both images and spirits. This is consistent with a structure where the word meaning 'reverence' is repeated with both 'images' and 'spirits'. If so, *þe* must mean 'reverence', *dab* must mean 'spirits', and *mlom* must mean 'images'. Note that the instructions warn that the order of words in the translations may not line up straightforwardly with the order of words in the Hmong expressions; this is a clear example where this is the case.

This kind of expression, where an ABAC pattern in Hmong lines up with an English translation of X and Y (or something similar in meaning) is crucial to solving this problem. As the techniques illustrated above are applied systematically to the remaining items, it becomes apparent that all of the three English sentences that are to be translated into Hmong must correspond to Hmong expressions of this type. An interesting property of these expressions is that they are *coordinate*. Coordinate expressions include phrases like *dogs and cats* and *red or blue*. The fact that they are coordinate means that they can be reversed without changing their meaning (as in *cats and dogs* and *blue or red*). As implied by the translation exercise, Hmong coordinate expressions are not reversible. The deepest puzzle of this problem is finding what determines the order of the parts in a coordinate expression like *þe dab þe mlom*. To do this, it is necessary to collect all of the coordinate expressions and explore hypotheses regarding what best predicts their ordering. A careful examination reveals that the tone, written as the final consonant in a word, is the best predictor of what order the parts of a Hmong coordinate expression take. For example, words with the -b tone occur before words with the -g tone but words with the -j tone occur before words with the -b tone. Note that it is not necessary, or even helpful, to know what these tones actually sound like; the crucial fact is the logical relationship among the tones.
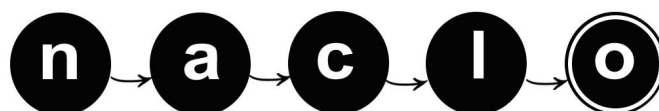
(P) The Old Man the Boats (1/2)

P1. [[The old] [train (verb) the young]] -> [[The old train (noun)] [broke down.]]
In *The old train the young*, *train* is a *verb* and it groups with the verb phrase *train the young*. In *The old train broke down*, *train* is a noun, which groups with the noun phrase *the old train*.

P2. Three solutions were accepted:

[[The thief [seized (passive verb) by the police]] [turned out to be our cousin]] ->
[[The thief] [seized (active verb) the laptop]]

In *The theif seized by the police turned out to be our cousin*, *seized* is part of a phrase (*seized by the police*) that modifies *thief*. Notice that the thief was seized. The thief did not do the seizing. The main verb of the sentence is *turned out*. The thief who was seized did the turning out. In *The thief seized the laptop*, *seized* is the main verb of the sentence. The thief did the seizing.

[[The thief [seized (passive verb) by the police]] [turned out to be our cousin]] ->
[[The thief [seized (passive verb) by [the police [turned traitor]]]]] [was our cousin]] ->

We accepted this answer because *turned* modifies *the police*. The thief isn't doing any kind of turning action. The police are the ones who turned.

[[The thief [seized (passive verb) by the police]] [turned out to be our cousin]] ->
[[The thief [seized (passive verb) by [the police station]]] [turned out to be our cousin]] ->

We accepted this answer because the police are no longer the ones doing the seizing. *The police* is simply a modifier of *station*. Note that the meaning of the word by has changed.

P3. [[I] [convinced [her children] [to do their homework]]] ->
[[I] [convinced [her] [(that) children do their homework]]]

In *I convinced her children to do their homework*, *her* modifies *children*. I convinced the children, not her. In *I convinced her (that) children do their homework*, I convinced her.

**P4.** [[The man who whistles] [tunes (verb) pianos]] ->
[[The man who whistles tunes (noun)] [also likes to sing]]

In this example, the local ambiguity is at the word *tunes*. In *The man who whistles tunes pianos*, *tunes* is a verb and the man does the tuning. In *The man who whistles tunes also likes to sing*, *tunes* is a noun and the tunes are being whistled.

P5. [[The cotton clothing (noun)] [is drying in the sun]] ->
[[The cotton [clothing (verb) the doll]] [is drying in the sun]] or
[[The cotton [clothing (noun) is made of]] [is drying in the sun]]
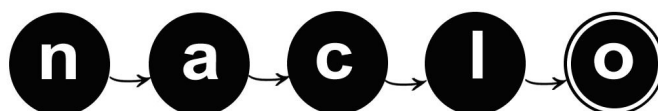
In *The cotton clothing is drying in the sun*, clothing is drying in the sun. In *The cotton clothing the doll is drying in the sun* and *The cotton clothing is made of is drying in the sun*, cotton is drying in the sun. *Clothing* is a modifier of *cotton*.